1-2015

Discrimination of Mild Cognitive Impairment and Alzheimer's Disease Using Transfer Entropy Measures of Scalp EEG

Joseph McBride
University of Tennessee

Xiaopeng Zhao
University of Tennessee

Nancy Munro
Oak Ridge National Laboratory

Gregory Jicha
University of Kentucky, gregory.jicha@uky.edu

Charles Smith
University of Kentucky, csmith@mri.uky.edu

See next page for additional authors

Follow this and additional works at: https://uknowledge.uky.edu/sbcoa_facpub

Part of the Family, Life Course, and Society Commons, and the Geriatrics Commons
Right click to open a feedback form in a new tab to let us know how this document benefits you.

Discrimination of Mild Cognitive Impairment and Alzheimer's Disease Using Transfer Entropy Measures of Scalp EEG

Notes/Citation Information

Authors

Joseph McBride, Xiaopeng Zhao, Nancy Munro, Gregory Jicha, Charles Smith, and Yang Jiang

Discrimination of Mild Cognitive Impairment and Alzheimer's Disease Using Transfer Entropy Measures of Scalp EEG

Joseph McBride, PhD[1]; Xiaopeng Zhao, PhD[1*]; Nancy Munro, PhD[2];
Gregory Jicha, MD, PhD[3,4]; Charles Smith, MD[3,4] and Yang Jiang, PhD[3,5]

[1]*Department of Mechanical, Aerospace, and Biomedical Engineering, University of Tennessee, Knoxville, TN, USA*
[2]*Oak Ridge Nation Laboratory, Oak Ridge, TN, USA*
[3]*Sanders-Brown Center on Aging,* [4]*Department of Neurology,* [5]*Department of Behavioral Science, University of Kentucky College of Medicine, Lexington, KY, USA*

ABSTRACT

Mild cognitive impairment (MCI) is a neurological condition related to early stages of dementia including Alzheimer's disease (AD). This study investigates the potential of measures of transfer entropy in scalp EEG for effectively discriminating between normal aging, MCI, and AD participants. Resting EEG records from 48 age-matched participants (mean age 75.7 years)—15 normal controls, 16 MCI, and 17 early AD—are examined. The mean temporal delays corresponding to peaks in inter-regional transfer entropy are computed and used as features to discriminate between the three groups of participants. Three-way classification schemes based on binary support vector machine models demonstrate overall discrimination accuracies of 91.7—93.8%, depending on the protocol condition. These results demonstrate the potential for EEG transfer entropy measures as biomarkers in identifying early MCI and AD. Moreover, the analyses based on short data segments (two minutes) render the method practical for a primary care setting.

Keywords: early Alzheimer's disease, mild cognitive impairment, EEG-based diagnosis, transfer entropy

*Corresponding author: Xiaopeng Zhao, Department of Mechanical, Aerospace, and Biomedical Engineering, National Institute of Mathematical and Biological Synthesis, University of Tennessee, Knoxville, TN 37996, USA. Email: xzhao9@utk.edu. Other authors: jmcbrid4@utk.edu; nbmunroconsulting@comcast.net; gajich2@email.uky.edu; csmith@mri.uky.edu; yjiang@uky.edu.

1. INTRODUCTION

Mild cognitive impairment (MCI) is a memory and cognition disruption associated with old age and a departure from normal aging. Amnestic MCI is often an early stage of dementia such as Alzheimer's disease (AD) [1, 2]. Identifying MCI state enables early intervention. Development of a simple electroencephalogram (EEG)-based screening method for MCI for use in the primary care setting would be a valuable tool in the research and treatment of both disorders.

Transfer entropy is an information theoretic measure that quantifies the statistical coherence between systems evolving time [3, 4]. While standard time-delayed mutual information fails to distinguish information that is actually exchanged from shared information due to common history and input signals, transfer entropy is able to effectively distinguish driving and responding elements and to detect asymmetry in the interaction of subsystems [3]. The fact that it is non-symmetric enables one to infer the direction of information flow. Transfer entropy reduces to Granger causality for processes in which the output variable depends linearly on its own previous values (i.e., vector auto-regressive processes) and Gaussian variables [5]. Transfer entropy has been applied to many different fields, including neuroscience, systems biology, bioinformatics, environmental sciences, climatology, engineering, finance, astronomy, Earth and space sciences, and astronomy. Transfer entropy has been used for estimation of functional connectivity of neurons [6, 7] and social influence in social networks. Recent studies indicate that cognitive declines in MCI and AD may manifest as reduced complexity and perturbations in EEG synchrony [8-11]. Much work has been done to show the group differences between different cognitive statuses. However, group differences often are not sufficient to establish the initial evaluation of subjects with cognitive impairment in the routine clinical practice [12]. This work aims to develop easy neural indicators to diagnose and predict cognitive decline pre-clinically, such as among people with subjective memory complaints.

While various studies have investigated applications of entropy and complexity measures of scalp EEG for discrimination of cognitive deficits [8, 13-18], the investigation on transfer entropy of EEG has been limited. Based on the hypothesis that the cognitive disruption of AD or MCI may affect the information exchange between two areas of the brain, this work aims to explore the application of transfer entropy of scalp EEG to diagnose AD and MCI. In this study, inter-regional EEG dynamics in MCI and AD during rest and a simple counting task is examined. Transfer entropy-based measures are studied as features to discriminate cognitive impairment of MCI and AD. To the authors' best knowledge, this article is the first study on transfer entropy analysis of scalp EEG for AD and MCI participants.

2. METHODS
2.1. Data

The EEG data used in this study were collected in the Behavioral Science Department and Sanders-Brown Center on Aging at the University of Kentucky (UK) College of Medicine. Participants between the ages of 60 and 90 years were recruited from a study cohort of cognitively normal older adults followed by the Alzheimer's Disease Center

(ADC) of the UK College of Medicine [19]. MCI patients were recruited from the Memory Disorders Clinic of the ADC. Normal older participants are screened regularly and when screenings indicate possible cognitive decline, they are referred to the ADC's Research Memory Disorders Clinic. MCI and AD participants were diagnosed and recruited by cognitive neurologists Drs. C. Smith and G. Jicha at the UK ADC Clinical Core and from its Research Memory Disorders Clinic. All participants provided written informed consent before participation. This study was approved by the Institutional Research Board (IRB) of the University of Kentucky. The database includes 15 NC, 16 MCI, and 17 AD participants. The mean (standard deviation) ages for NC, MCI, and AD participants are 75.7 years (5.5 years), 74.6 years (9.0 years), and 76.7 years (5.2 years), respectively.

All MCI participants belonged to the amnestic subtype. In addition, a few of the MCI participants also presented with executive dysfunction. Differences between single- and multiple-domain MCI subtypes are not the focus of the current study. A list of neurological assessments used to make MCI/AD diagnoses is provided in Table 1. Means (standard deviation) for Mini-Mental State Examination (MMSE) scores were 29.43 (0.73), 27.40 (2.00), and 24.26 (2.42) for the NC, MCI, and AD groups,

Table 1. UK-ADC uniform research battery of cognitive tests and other evaluations

General Cognitive Measures	**Baseline Only**
MMSE; Kokmen Test	National Adult Reading Test*
Clinical Dementia Rating (CDR)	
Attention/Executive Domain Measures	**Medical Evaluation**
Trail Making Tests A & B	Physical exam*
WAIS-R Digit Span & Digit Symbol	Neurological exam*
	Medical history*
Memory Domain Measures	**Medications**
WMS Logical Memory I & II	Nutritional supplements
California Verbal Learning Test*	Food Frequency Questionnaire (FFQ)
Language Domain Measures	**Psychiatric Evaluation**
Controlled Oral Word Association Test (COWAT)*	Neuropsychiatric Inventory Questionnaire (NPI-Q)
Animal & Vegetable Fluency	Geriatric Depression Scale (GDS)
Boston Naming	
Visual/Spatial Domain Measures	**Functional Ability Measures**
CERAD Figures	Functional Assessment Questionnaire (FAQ)
	SF-36*
	ADCS-ADL*

*Additional test measures or expanded content beyond Uniform Data Set (UDS) core measures

respectively. MCI and early AD participants' EEG data were recorded as soon as possible after diagnoses. Patients with stroke or seizure history were not included.

Participants were connected to 64- or 32-channel EEG caps using a Neuroscan II system (10–20 montage). In either case, only the 32 common channels were recorded. EEG data were recorded under a protocol using three different non-memory-task conditions. These included (1) resting with eyes open for 5 min (REO condition), (2) resting with eyes closed while counting backwards by ones for 10 min while tapping a finger (counting task), and (3) resting with eyes closed for 10 min (REC condition). The EEG recording was performed without interruption at the same appointment for each subject. EEG data were acquired at 500 Hz. The 32 EEG channels included 2 ocular channels that were used to determine the dominant eye blink frequency. Notch filters were used to remove dominant eye blink frequencies and to remove 60 Hz frequencies, which may have been amplified by background electronic devices. A simple 2nd order Butterworth filter was used to attenuate frequencies greater than 200 Hz.

2.2. Transfer Entropy

Information can be defined mathematically as that which decreases uncertainty [20]. Entropy, being a measure of the uncertainty in a system, can therefore be used to mathematically quantify the basic amount of information in time series [20]. For example, the average number of bits required to optimally encode independent samples of a discrete variable x, which follows a probability distribution function p_x, is given by the classical definition for Shannon entropy:

$$S = -\sum_{i=1}^{N} p_x(x_i) \log_2 \left[p_x(x_i) \right], \tag{1}$$

where the sum extends over all possible states x_i ($i = 1, ..., N$) that the random variable x can assume. The base of the logarithm depends only on the units used for measuring information; thus, it will be dropped in further equations. If a different distribution is used, there will be an excess number of bits and the information will not be optimally encoded [3]. The excess number of bits that will be encoded if a different distribution function q is used is given by the Kullback entropy [21]:

$$K = \sum_{i=1}^{N} p_x(x_i) \log \left[\frac{p_x(x_i)}{q(x_i)} \right]. \tag{2}$$

The mutual information of two random variables x and y with joint probability distribution function $p_{x,y}$ can be viewed as the excess amount of code produced when the two variables are assumed to be independent. That is, assuming $p_{x,y}(x_i, y_j) = p_x(x_i)p_y(y_j)$. Mutual information M_{xy} between the variables x and y is thus the Kullback entropy given the assumption of independence between the two variables:

$$M_{xy} = \sum_{i=1}^{N} \sum_{j=1}^{M} p_{x,y}(x_i, y_j) \log \left[\frac{p_{x,y}(x_i, y_j)}{p_x(x_i)p_y(y_j)} \right], \tag{3}$$

where the random variable y can assume states y_i ($j = 1, ..., M$). Mutual information thus provides an intuitive means for quantifying the deviation from independence of two random variables. Note that mutual information M_{xy} is symmetric under the exchange of x and y and therefore does not contain any information regarding the directional transfer of information (which may relate to causal effect) [22].

Suppose that the variables x and y are generated by two separate but coupled systems. An entropy rate S_1 can be defined as the amount of additional information required to represent the current value x_n of the random variable x given the value of both variables at an observation d steps in the past:

$$S_1 = -\sum_{x_n}\sum_{x_{n-d}}\sum_{y_{n-d}} p_{x_n,x_{n-d},y_{n-d}}\left(x_n,x_{n-d},y_{n-d}\right)\log\left[p_{x_n|x_{n-d},y_{n-d}}\left(x_n|x_{n-d},y_{n-d}\right)\right], \qquad (4)$$

where $p_{x_n,x_{n-d},y_{n-d}}$ is the joint probability function for the observations x_n, x_{n-d}, y_{n-d}, and y_{n-d}, and $p_{x_n|x_{n-d},y_{n-d}}$ is the conditional probability function for x_n given x_{n-d} and y_{n-d}. Further suppose that the value of the current observation x_n of the random variable x is independent of the value of the random variable y observed d steps in the past (y_{n-d}). The entropy rate in this case would then reduce to the following:

$$(5)$$

$$S_2 = -\sum_{x_n}\sum_{x_{n-d}}\sum_{y_{n-d}} p_{x_n,x_{n-d},y_{n-d}}\left(x_n,x_{n-d},y_{n-d}\right)\log\left[p_{x_n|x_{n-d}}\left(x_n|x_{n-d}\right)\right],$$

where $p_{x_n|x_{n-d}}$ is the conditional probability function for x_n given x_{n-d}. The incorrectness of this assumption can then be express as the difference $S_1 - S_2$ between the two entropy rates presented in Equations (4) and (5). This difference in entropy is then termed transfer entropy $T^d_{y\rightarrow x}$ [3]:

$$(6)$$

$$T^d_{y\rightarrow x} = \sum_{x_n}\sum_{x_{n-d}}\sum_{y_{n-d}} p_{x_n,x_{n-d},y_{n-d}}\left(x_n,x_{n-d},y_{n-d}\right)\ \log\left[\frac{p_{x_n|x_{n-d},y_{n-d}}\left(x_n|x_{n-d},y_{n-d}\right)}{p_{x_n|x_{n-d}}\left(x_n|x_{n-d}\right)}\right]$$

Substituting the definitions for the conditional probabilities given in Equations (7) and (8), Equation (6) reduces to Equation (9):

$$p_{x_n|x_{n-d},y_{n-d}}\left(x_n|x_{n-d},y_{n-d}\right) = \frac{p_{x_n,x_{n-d},y_{n-d}}\left(x_n,x_{n-d},y_{n-d}\right)}{p_{x_{n-d},y_{n-d}}\left(x_{n-d},y_{n-d}\right)}, \qquad (7)$$

$$p_{x_n|x_{n-d}}\left(x_n|x_{n-d}\right) = \frac{p_{x_n,x_{n-d}}\left(x_n,x_{n-d}\right)}{p_x\left(x_n\right)}, \qquad (8)$$

$$T_{y \to x}^{d} = \sum_{x_n} \sum_{x_{n-d}} \sum_{y_{n-d}} p_{x_n, x_{n-d}, y_{n-d}} \left(x_n, x_{n-d}, y_{n-d} \right)$$

$$\log \left[\frac{p_{x_n, x_{n-d}, y_{n-d}} \left(x_n, x_{n-d}, y_{n-d} \right) p_x \left(x_n \right)}{p_{x_{n-d}, y_{n-d}} \left(x_n, y_{n-d} \right) p_{x_n, x_{n-d}} \left(x_n, x_{n-d} \right)} \right], \tag{9}$$

where $p_{x_{n-d}, y_{n-d}}$ is the joint probability of variables x_{n-d} given y_{n-d}, and $p_{x_{n-d}, y_{n-d}}$ is the joint probability of variables x_n given x_{n-d}. The transfer entropy $T_{y \to x}^{d}$ is explicitly not symmetric since it measures the degree of dependence of x on y and not vice versa. Thus, $T_{y \to x}^{d} \neq T_{x \to y}^{d}$. Note that the definition of transfer entropy here follows the definition in [3]. Recently, Wibral et al. proposed an improved definition of transfer entropy to better measure the delay effect [23]. The differences between the two definitions shall be investigated in future research.

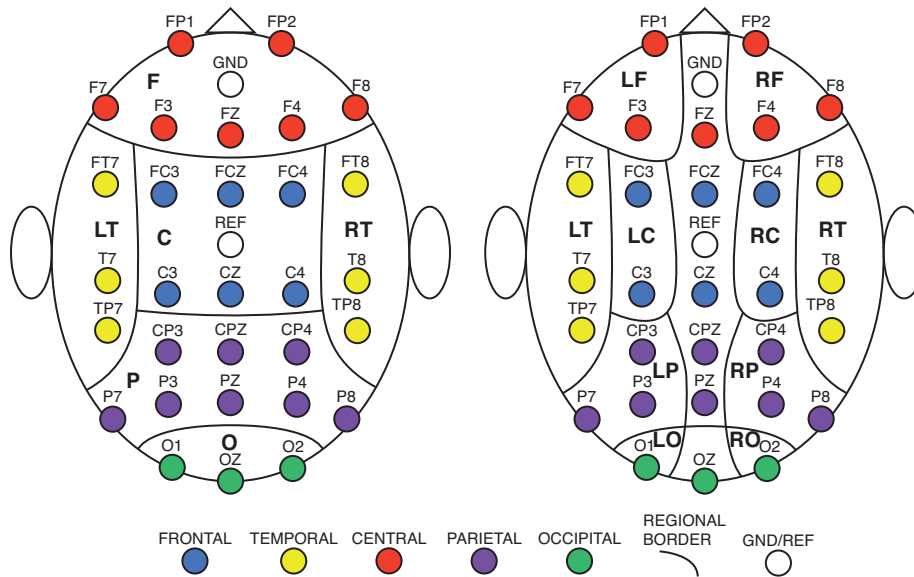## 2.3. Peak Inter-regional Transfer Entropy Delays (PITEDs)

The 30 channels were grouped into 14 scalp regions based on their arrangement and location on the scalp. The regions included: (1) left frontal (LF); (2) right frontal (RF); (3) frontal (F = LF + RF + channel FZ); (4) left temporal (LT); (5) right temporal (RT); (6) left central (LC); (7) right central (RC); (8) central (C = LC + RC + channels FCZ and CZ); (9) left parietal (LP); (10) right parietal (RP); (11) parietal (P = LP + RP + channels CPZ and PZ); (12) left occipital (LO); (13) right occipital (RO); and (14) occipital (O = LO + RO + channel OZ). Note that left and right regions do not include central line channels; see Figure 1 for regional boundaries.

Peak values and peak delays are the most important characteristics of transfer entropy. Since channels have different magnitudes, peak values may not accurately reflect the effect of the measurements of information exchange between regions. We chose to use peak delay as the features since they are not influenced by magnitudes of the channels. First, transfer entropies were computed for each directional, pairwise combination of 30 channels (2 ocular channels excluded) for each protocol condition. The first two minutes of data for each protocol condition was used. Entropies were computed for delays of 0.002 through 1 second in 0.002-second steps. The delay at which the transfer entropy was greatest in magnitude was noted as the peak delay. Then, the inter-region delay for a directional pairing of regions (for example, region X → region Y) was computed as the average of all the delays for each channel in region X to each channel in region Y; see Equation (10),

$$PITED_{x \to y} = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} d_{i \to j}^{peak} \tag{10}$$

where $d_{i \to j}^{peak}$ is the peak delay for channel i in region X to channel j in region Y, N_x is the number of channels in region X, and N_y is the number of channels in region Y. Note that inter-regional delays were defined for two different regions and they
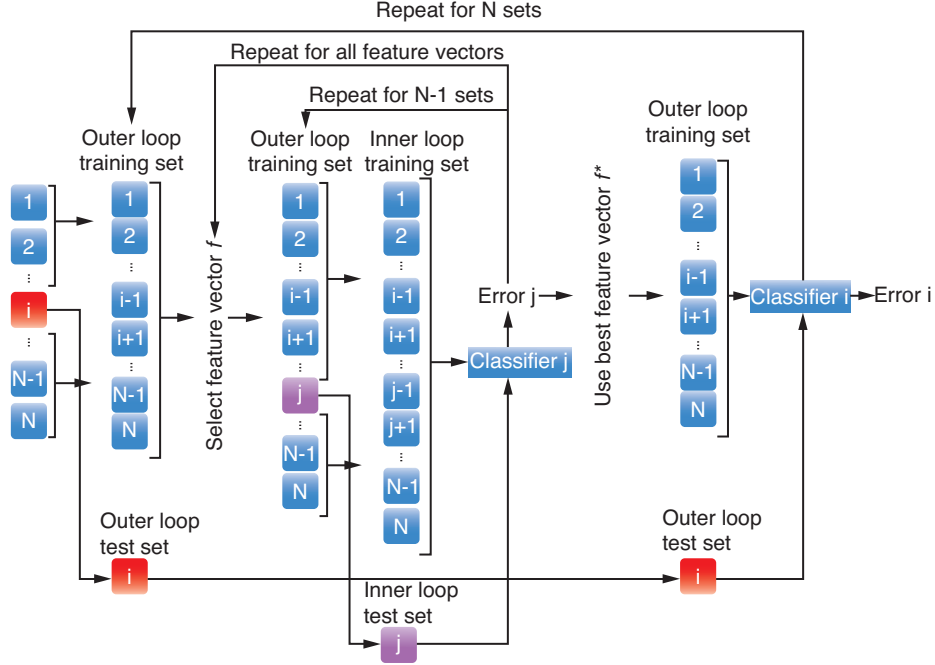
Figure 1.   Regional and Subregional Boundaries for Electrodes. Left: major regions; right: subregions. LF = left frontal; RF = right frontal; F = frontal; LT = left temporal; RT = right temporal; LC = left central; RC = right central; C = central; LP = left parietal; RP = right parietal; P = parietal; O = occipital. Note that central line channels (those with Z in the designation) are excluded from L/R subregions.

were not computed between a major region (C, F, O, P, or R) and any of its sub-regions. For example, the peak inter-regional delays for F→ LF or LF→ F were not computed. Thus, a total of 166 directional, mean peak inter-regional transfer entropy delays (PITEDs) were determined for each protocol condition.

2.4. Classification

For each protocol condition (REO, REC, or counting), we performed two types of classifications: first, binary classifications were conducted to discriminate (1) MCI vs. NC, (2) AD vs. NC, and (3) MCI vs. AD; and then a three-way classification was performed based on the binary classification results to determine which of the three groups a given record belonged to. The three-way classification scheme allows for the differentiation of the three groups without a priori knowledge that individuals fall into two of the three groups. Note that results of all classifications (both binary and three-way) were based on leave-one-out cross-validation (LOOCV) to avoid overfitting. For binary classifications, we used the support vector machine (SVM) functions in MATLAB [24] with quadratic kernel functions. A three-way classification scheme is constructed by combining the outcomes of the binary classifiers using the pairwise coupling approach proposed by Hastie and Tibshirani

Figure 2.   Schematic feature selection process.

[25]. For a given record, binary SVM classifiers (i.e., MCI vs. NC, AD vs. NC, and MCI vs. AD) are trained using all other available records and then applied to the given record. If two out of three of the SVM binary classifiers classify a record as belonging to class i, then the final decision of the three-way classifier is to classify the record as belonging to class i. Otherwise, the probability that a record belongs to each class, P_i, i = 1, 2, 3, is then estimated via pairwise coupling and then the final decision of the three-way classifier is to choose the class corresponding to the largest probability, $\text{argmax}_i(P_i)$.

2.5. Feature Selection
Feature selection was performed for each of three binary discrimination problems [(1) MCI vs. NC, (2) AD vs. NC, and (3) MCI vs. AD] for each protocol condition; see the schematic process in Figure 2. We used the SVM functions in MATLAB [24] as the binary classifier. Quadratic kernel functions were used in all discriminations and the cost coefficient was held constant at unity. Nested LOOCV loops were used to avoid overfitting while suggesting and testing different combinations of PITEDs as features [26, 27]. The inner loop generated a list of suggested combinations via a forward, high-score, features selection method where combinations were scored using LOOCV accuracy of SVM model predictions from

a smaller, randomized, subset of records [26]. The outer loop determined the LOOCV accuracy of combinations suggested by the inner loop for *all* available records. The discriminatory power of individual PITEDs was then assessed based on how often they appeared in the 100 best performing combinations tested in the outer loop simulation.

2.6. Statistical Significance

Monte Carlo permutation testing was used to assess the statistical significance of the LOOCV accuracies of the binary classifiers. Specifically, a random sample of 10,000 permutations of shuffled labels indicating groups (NC, MCI, or AD) was used to estimate a 95% confidence interval for the probability that the leave-one-out cross-validation accuracies obtained were due to chance. The p-values presented were determined using this method [28].

3. RESULTS
3.1. Binary Discrimination Results

Two-sample Student's *t*-distribution tests (unequal variance) are performed on group means of PITEDs selected as features for binary classifiers in order to determine if observed differences are significant enough to infer a linear separability at the population level. Such an inference would require large differences between group means and small variation within groups. It should be noted that inferences are dependent on the assumption of representative samples.

3.1.1. MCI vs. NC

Results for LOOCV accuracies for each binary classifier are presented in Table 2 along with the selected features. Accuracies presented are obtained using selected features. For the REO condition, MCI participants demonstrate a lower PITED for occipital region to right frontal region coupled with an increased PITED from the frontal region to the left occipital region, compared to normal controls. An increased central-to-frontal PITED is also observed for MCI participants. Most of the selected features for the counting task include PITEDs of directed information transfer to the occipital region. For the REC condition, the pathway of right frontal to right temporal to right parietal to occipital appears discriminate MCI participants. Ninety-five percent confidence intervals for the corresponding p-values are determined via Monte Carlo permutation testing. LOOCV accuracies of 93.6% (p < 0.0032), 90.3% (p < 0.0182), and 87.1% (p < 0.0321) are achieved for MCI vs. NC discrimination for the REO condition, counting task, and REC condition, respectively.

3.1.2. AD vs. NC

Compared to normal controls, AD participants demonstrate a significantly greater left temporal-to-frontal PITED coupled with a decrease in frontal-to-left temporal PITED during the REO condition. A greater right occipital-to-left central PITED is also observed for AD participants. A LOOCV accuracy of 93.8% (p < 0.0043) is achieved for the REO condition based on these differences. During the counting task, AD

Table 2. Group comparisons of selected features for binary classifiers

Condition	MCI vs. NC		AD vs. NC		AD vs. MCI	
	Selected Features	Group Means	Selected Features	Group Means	Selected Features	Group Means
REO	O → RF	MCI<NC	LT → F	AD>NC	LT → LF	MCI<AD***
	F → LO	MCI>NC	RO → LC	AD>NC	RO → LC	MCI>AD
	C → F	MCI>NC	F → LT	AD<NC	P → LO	MCI<AD
	acc. (sens., spec.): 93.6% (100%, 86.7%) 95% CI for p-value: (0.0001, 0.0032)		acc. (sens., spec.): 93.8% (100%, 86.7%) 95% CI for p-value: (0.0031, 0.0043)		acc. (sens., spec.): 90.9% (93.8%, 88.2%) 95% CI for p-value: (0.0097, 0.0118)	
Counting Task	RF → LP	MCI>NC	RP → LO	AD>NC***	RP → LO	MCI<AD**
	P → LT	MCI<NC	RO → LO	AD>NC*	RP → O	MCI<AD***
	LP → LO	MCI<NC	LF → LO	AD>NC*	LF → LT	MCI<AD*
	acc. (sens., spec.): 90.3% (93.8%, 86.7%) 95% CI for p-value: (0.00138, 0.0182)		acc. (sens., spec.): 90.6& (82.4%, 100%) 95% CI for p-value: (0.0020, 0.0110)		acc. (sens., spec.): 90.9% (88.2%, 93.8%) 95% CI for p-value: (0.0032, 0.0195)	
REC	LO → RF	MCI>NC**	LT → RC	AD<NC	RF → RT	MCI<AD*
	C → RT	MCI<NC**	O → LF	AD>NC	LC → O	MCI<AD
	RT → RP	MCI>NC	LO → LT	AD<NC	RP → O	MCI<AD
	acc. (sens., spec.): 87.1% (87.5%, 86.7%) 95% CI for p-value: (0.0057, 0.0321)		acc. (sens., spec.): 87.5% (82.4%, 93.3%) 95% CI for p-value: (0.0211, 0.0324)		acc. (sens., spec.): 81.8% (100%, 64.7%) 95% CI for p-value: (0.0162, 0.0373)	

* p < 0.05 ** p < 0.01, ***p < 0.001. Feature designations are preceded by regional indices: LF = left frontal; RF = right frontal; F = frontal; LT = left temporal; RT = right temporal; LC = left central; RC = right central; C = central; LP = left parietal; RP = right parietal; P = parietal; LO = left occipital; RO = right occipital; O = occipital; see Figure 1 for regional boundaries. REO = resting with eyes open; REC = resting with eyes closed. acc. = accuracy, sens. = sensitivity, and spec. = specificity.

participants demonstrate significantly increased PITEDs for right parietal-to-left occipital, left frontal-to-left occipital, and right occipital-to-left occipital. These observations allow for a LOOCV accuracy of 90.6% (p < 0.0110) for AD vs. NC discrimination during the counting task. When resting with eyes closed, AD participants demonstrated an increase in occipital-to-left frontal PITED and a decrease in left temporal-to-right central PITED coupled with a decrease in left occipital-to-left temporal PTITED. An accuracy of 87.5% (p < 0.0324) is achieved for the REC condition.

3.1.3. *MCI vs. AD*

For the REO condition, AD participants demonstrate a significantly greater left temporal-to-left frontal PITED compared to MCI participants. This appears to closely follow the observation of increased left temporal-to-frontal PITED in AD participants compared to normal controls. Interestingly, during the REO condition, the right occipital-to-left central PITED is greater for AD participants compared to NC participants but lower for AD participants compared to MCI participants. There is also an observed increase in parietal-to-left occipital PITED for AD participants. A LOOCV accuracy of 90.9% ($p < 0.0118$) for MCI vs. AD discrimination is achieved based on these observations for the eyes open resting condition. MCI vs. AD LOOCV accuracies of 84.5% ($p < 0.0195$) and 81.8% ($p < 0.0373$) are achieved for the counting task and REC condition, respectively. During the counting task, AD participants demonstrate significantly increased right occipital-to-left central PITED compared to MCI participants and NC participants. AD participants also have significantly higher right parietal-to-occipital and left frontal-to-left temporal PITEDs. While resting with eyes closed, AD participants have higher right frontal-to-right temporal, left central-to-occipital, and right parietal-to-occipital PITEDs.

3.2. Three-way Classification Results

Table 3 is a contingency table that shows the performance of the three-way classification for the REO condition. One participant from each group is misclassified using data from the REO condition: one NC participant is misclassified as AD, one MCI participant is misclassified as NC, and one AD participant is misclassified as MCI. Overall, the accuracy is 93.8%.

Table 4 shows a summary of the three-way classification results for the counting task. All NC participants are correctly classified. Two MCI participants are misclassified as NC participants. Two AD participants are also misclassified, one as MCI and one as NC. Thus, 100% of those predicted as AD are AD, 93.3% of those predicted to be MCI are MCI, and 83.3% of those predicted to belong to the NC group are actually NC participants. The resulting overall accuracy is 91.7%.

Results for the REC condition are very similar to those of the REO condition and are presented in Table 5. Comparing Tables 3 and 5, the only differences are that an

Table 3. Contingency table of 3-way classification results for REO condition using PITEDs.

		Predicted Classes			
		NC	MCI	AD	
True Classes	NC	14	0	1	93.3%
	MCI	1	15	0	93.8%
	AD	0	1	16	94.1%
		93.3%	93.8%	94.1%	Overall Acc.: 93.8%

Table 4. Contingency Table of 3-way classification results for counting task using PITEDs.

		Predicted Classes			
		NC	MCI	AD	
True Classes	NC	15	0	0	100%
	MCI	2	14	0	87.5%
	AD	1	1	15	88.2%
		83.3%	93.3%	100%	Overall Acc.: 91.7%

Table 5. Contingency table of 3-way classification results for REC condition using PITEDs.

		Predicted Classes			
		NC	MCI	AD	
True Classes	NC	14	1	0	93.3%
	MCI	1	14	1	87.5%
	AD	0	1	16	94.1%
		93.3%	87.5%	94.1%	Overall Acc.: 91.7%

NC participant is misclassified as MCI instead of being misclassified as AD and an MCI participant is also misclassified as NC. The overall accuracy for the REC condition is the same as for the counting task at 91.7%.

4. DISCUSSION
4.1. Contextualization of Results
In this study, average delays corresponding to maximized transfer entropy between EEG channels of different scalp regions are examined in normal aging, MCI, and early AD participants. We have analyzed EEG data collected using three different EEG protocols: *resting eyes open condition, resting eyes closed condition,* and a *simple cognitive task* of *counting backwards by ones*. Accuracies of 93.6-97.0% (using the three EEG protocol conditions) are achieved for binary classifications between the three groups using features from average peak inter-regional transfer entropy delays (PITEDs). A three-way classification scheme is also derived from the binary classifiers based on pairwise coupling, with overall accuracies of 91.7-93.8% (using the three EEG protocol conditions). The resting-state protocols reflect default mode networking activity in the brain and have shown promise in clinical application of AD [29].

Counting and other cognitive protocols reflect task related network in the brain. Both method have demonstrated the EEG biomarkers' are associated with AD pathology [30].

As observed in previous studies [31-34], features that successfully discriminate between AD and NC may not be able to discriminate between MCI and NC or between AD and MCI. As such, previous studies often investigated the three binary classification problems using different sets of features for each classification. Moreover, none of the aforementioned studies reported three-way classification among NC, MCI, and AD participants. A three-way classification model was constructed by combining the binary classifiers in [18], where the results demonstrate overall discrimination accuracies of 83.3%, 85.4%, and 79.2% for resting eyes open, counting eyes closed, and resting eyes closed protocols, respectively. Therefore, the encouraging results of three-way classifications in this work indicate the robustness of the transfer entropy measures. The three-way classification scheme may be used for differentiation of the three groups without a priori knowledge that individuals fall into two of the three groups. In contrast to the previous studies on differences between the cognitive groups, this work classifies participants at the individual level, which is the case when a patient comes into a doctor's office. Moreover, the individualized analysis also makes it easy to test these EEG markers against known biomarkers and known cognitive tests.

The most significant results indicate that the transmission of electrophysiological activity at the parietal electrode site to early visual cortex is delayed in AD patients compared to those in MCI or in NC. The disruption of white matter connections has been linked to delayed reaction times or processing speed in the brain [35]. Recent evidence has shown that white matter lesion and Alzheimer's Tau pathology at the parietal lobe both contribute to development of AD [36]. The current entropy results of delayed parietal activity in AD are consistent with AD pathology. In addition, it is also well-known that medial temporal cortex, important structure for memory, is among the earliest to be affected in AD pathology. Our second most significant result showed delayed activity from left temporal cortex to left frontal in AD compared to that in NC during resting eyes open state. This result indicates the deficits in the default-mode network between left temporal-frontal communication. With further validation, these parietal and temporal activity-related EEG indicators are likely promising markers with clinical implication of AD progression. In addition, individual EEG indicators among normal older adults may be used as a predictor for risk factor of AD.

4.2. Limitations and Future Work
The current method cannot be readily applied to the clinical setting since it requires 30 EEG channels, making setup time an issue. Fewer electrodes would be ideal for more convenient application. Another shortcoming of the current study is the relatively small sample size. Future work should examine the developed EEG indicators against known biomarkers (e.g., cerebrospinal fluid proteins) in a larger sample size. Future work should also investigate whether the EEG indicators bear longitudinal changes over relatively short time periods at the individual level. Future work may also consider data fusion of the features from different conditions to allow synergies between the features.

5. CONCLUSIONS

Analysis of EEG inter-channel transfer entropy allows for observations of patterns in the complexity and synchrony of EEG data. Specifically, the features proposed here are intended to represent trends in the delay of the transfer of the information present in EEG signals between major regions of the scalp. The successful discrimination between the three groups of EEG records (NC, MCI, and AD) are the result of differences in the dynamics of the distribution of information in EEG voltages across the scalp in resting states and during a simple cognitive task. It is possible that the observed differences in these information dynamics may be influenced by alterations in the functional organization of the brain as a result of cognitive decline. The results of this pilot study suggest the potential for the use of features representing inter-regional EEG transfer entropy relationships as a means for objectively discriminating between normal older, MCI, and AD participants.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare.

REFERENCES

[1] Petersen, R.C., R. Doody, A. Kurz, R.C. Mohs, J.C. Morris, P.V. Rabins, K. Ritchie, M. Rossor, L. Thal, and B. Winblad. Current concepts in mild cognitive impairment. Arch. Neurol. 2001, 58: 1985–1992.

[2] Petersen, R.C. Mild cognitive impairment. New York, NY: Oxford Press, 2003.

[3] Schreiber, T. Measuring information transfer. Phys. Rev. Lett. 2000, 85: 461–464.

[4]   Hlaváčková-Schindler, K., M. Palus, M. Vejmelka, and J. Bhattacharya. Causality detection based on information-theoretic approaches in time series analysis. Phys. Rep. 2007, 441: 1–46.

[5]   Barnett, L. Granger causality and transfer entropy are equivalent for Gaussian variables, Phys. Rev. Lett., 2009, 103: 238701.

[6]   Lizier, J., J. Heinzle, A. Horstmann, J. Haynes, M. Prokopenko. Multivariate information-theoretic measures reveal directed information structure and task relevant changes in fMRI connectivity, J. Comput. Neurosci. 2011, 30: 85–107.

[7]   Vicente, R., M. Vibral, M. Lindner, and G. Pipa. Transfer entropy—a model-free measure of effective connectivity for the neurosciences, J. Comput. Neurosci. 2011, 30: 45–67.

[8]   Dauwels, J., F. Vialatte, A., and A. Cichocki. Diagnosis of Alzheimer's disease from EEG signals: where are we standing? Curr. Alzheimer Res. 2010, 7(6): 487–505.

[9]   Stam, C.J., Y. van der Made, Y.A. Pijnenburg, and P. Scheltens. EEG synchronization in mild cognitive impairment and Alzheimer's disease. Acta Neurol. Scand. 2003, 108: 90–96,.

[10]   Stam, C.J., T. Montex, B.F. Jones, S.A. Rombouts, Y. van der Made, Y.A. Pijnenburg, and P. Scheltens. Disturbed fluctuations of resting state EEG synchronization in Alzheimer's disease. Clin. Neurophysiol. 2005, 116: 708–715.

[11]   Stam, C.J, B.F. Jones, G. Nolte, M. Breakspear, and P. Scheltens. Small-world networks and functional connectivity in Alzheimer's disease. Cerebral Cortex. 2007, 17: 92–99.

[12]   Jelic, V. and J. Kowalski. Evidence-based evaluation of diagnostic accuracy of resting EEG in dementia and mild cognitive impairment. Clin. EEG Neurosci. 2009, 40: 129–142.

[13]   Sneddon, R. W. Shankle, J. Hara, A. Rodriguez, D. Horrman, and U. Saha. EEG detection of early Alzheimer's disease using psychophysical tasks. Clin. EEG Neurosci. 2005, 36(3): 141–150.

[14]   Zhao, P., P. Van-Eetvelt, C. Goh, N. Hudson, S. Wimalaratna, and E. Ifeachor. Characterization of EEGs in Alzheimer's disease using information theoretic methods. Proc. IEEE Eng. Med. Biol. Soc. 2007: 512705131.

[15]   McBride, J., X. Zhao, N. Munro, C. Smith, G. Jicha, and Y. Jiang. "Resting EEG discrimination of early stage Alzheimer's disease from normal aging using inter-channel coherence network graphs," Ann. Biomed. Eng., 2013, vol. 41, pp. 1233–42.

[16]   McBride J., X. Zhao, T. Nichols, V. Vagnini, N. Munro, D. Berry, and Y. Jiang. "Scalp EEG-based discrimination of cognitive deficits after traumatic brain injury using event-related Tsallis entropy analysis," IEEE Trans. Biomed. Eng., 2013, vol. 60, pp. 90–96.

[17]   McBride J., X. Zhao, N. Munro, G. Jicha, F. Schmitt, R. Kryscio, C. Smith, and Y. Jiang. "Sugihara causality analysis of scalp EEG for detection of early Alzheimer's disease," NeuroImage: Clinical, 2015, vol. 7, pp. 258-265.

[18]   McBride J., X. Zhao, N. Munro, C. Smith, G. Jicha, L. Hively, L. Broster, F. Schmitt, R Kryscio, and Y. Jiang. "Spectral and complexity analysis of scalp EEG characteristics for mild cognitive impairment and early Alzheimer's disease," Comput. Meth. Prog. Bio. 2014, vol. 114, pp. 153–163.

[19]   F. Schmitt, P. Nelson, E. Abner, S. Scheff, G. Jicha, C. Smith, G. Cooper, M. Mendiondo, D. Danner, L. van Eldik, A. Caban-Holt, M. Lovell, and R. Kryscio. "University of Kentucky Sanders-Brown healthy brain aging volunteers: donor characteristics, procedures and neuropathology," Curr. Alzheimer Res., 2012, vol. 9, pp. 724–733.

[20]   Sneddon, R. The Tsallis entropy of natural information. Physica A. 2007, 386(1): 101–118.

[21]   Kullback, S. Information Theory and Statistics. New York, NY: Wiley, 1959.

[22]   Lizier, J.T. and Prokopenko, M. Differentiating information transfer and causal effect, European Physical Journal B, 2010, 73:605–615.

[23]   Wibral M, Pampu N, Priesemann V, Siebenhühner F, Seiwert H, et al. Measuring Information-Transfer Delays. PLoS ONE 2013, 8(2): e55809.

[24]   MathWorks, "MATLAB." Available at mathworks.com/products/matlab. Accessed Feb. 2013.

[25]   Hastie, T. and R. Tibshirani. Classification by pairwise coupling. Ann. Statist. 1998, 26: 451–471.

[26]   Bishop, C. Neural Networks for Pattern Recognition, pp. 295-329, Oxford Univ. Press, 2008.

[27]   Nowotny T. "Two challenges of correct validation in pattern recognition," Frontiers in Robotics and AI, 2014, Vol. 1, Article 5, pages 1–6.

[28]   Nichols, T. and A. Holmes. Nonparametric permutation tests for functional neuroimaging: a primer with examples, Hum. Brain. Map. 2001, 15: 1–25.

[29]   *Vecchio F*, *Babiloni C*, *Lizio R*, *Fallani Fde V*, *Blinowska K*, *Verrienti G*, *Frisoni G*, *Rossini PM*. Resting state cortical EEG rhythms in Alzheimer's disease: toward EEG markers for clinical applications: a review.*Suppl Clin Neurophysiol*. 2013, 62:223–36.

[30]   Olichney, J. M., Pak, J., Salmon, D. P., Yang, J.-C., Gahagan, T., Nowacki, R., et al.. Abnormal P600 word repetition effect in elderly persons with preclinical Alzheimer's disease. *Cognitive Neuroscience*, 2013, *1*(1), 1–9.

[31]   Baker, M., K. Akrofi, R. Schiffer, and M.W. O'Boyle. EEG patterns in mild cognitive impairment (MCI) patients. Open Neuroimag., 2008 J. 2: 52–55.

[32]   Huang, C., L. Wahlund, T. Dierks, P. Julin, B. Winblad, V. Jelic. Discrimination of Alzheimer's disease and mild cognitive impairment by equivalent EEG sources: a cross-sectional and longitudinal study. Clin. Neurophysiol. 2000, 111: 1961–1967.

[33]   Iqbal, K, A.C. Alonso, S. Chen, M.O. Chohan, E. El-Akkad, C.X. Gong, S. Khatoon, B. Li, F. Liu, A. Rahman, H. Tanikuai, and I. Grundke-Iqbal. Tau pathology in Alzheimer's disease and other tauopathies. Biochim. Biophys. Acta. 2005, 1793: 198–210.

[34]   Rossini, P.M., C. Del Percio, P. Pasqualetti, E. Cassetta, G. Binetti, G. Dal Forno, F. Ferreri, G. Frisoni, P. Chiovenda, C. Miniussi, L. Paris, M. Tombini, F. Vecchio, and C. Babiloni. Conversion from mild cognitive impairment to Alzheimer's disease is predicted by source and coherence of brain electroencephalography rhythms. Neuroscience. 2006, 143: 793–803.

[35]   Gold, B, Powell, DK, Liang, X, Jiang, Y, & Hardy, PA, Speed of lexical decision correlates with diffusion anisotropy in left parietal and frontal white matter: evidence from diffusion tensor imaging, *Neuropsychologia*, 2007, 45, 2439–2446.

[36]   Hertze, J., Palmqvist, S., Minthon, L., & Hansson, O. Tau Pathology and Parietal White Matter Lesions Have Independent but Synergistic Effects on Early Development of Alzheimer's Disease. *Dementia and Geriatric Cognitive Disorders Extra*, 2013, *3*(1), 113–122.

[37]   Lake, D., J. Richmann, M. Griffin, and J. Moorman. Sample entropy analysis of neonatal heart rate variability. Am. J. Physiol. 2002, 283(3): R789-R797.